

Varsplic 設定方法 (Swiss Prot)

Swiss Prot は、ひとつのタンパク質に splice isoform、variant、conflict が存在する場合は、それらの情報をそのタンパク質のアノテーションとしてリファレンスファイル (DAT ファイル) にまとめています。従いまして、Swiss Prot にはスプライシングの内容が反映されたタンパク質エントリは含まれていません。

EBI (European Bioinformatics Institute) が提供する「`varsplc.pl`」スクリプトと「Swissknife」Perl ライブラリを利用すると、リファレンスファイルに記載されているスプライシング内容が反映されたタンパク質エントリを作成することができます。詳しくは、

<ftp://ftp.ebi.ac.uk/pub/software/swissprot/varsplc/README>

を参照してください。

Mascot Server が装備している配列データベース自動更新スクリプト、

`C:\inetpub\mascot\bin\db_update.pl`

と EBI が提供する `varsplc.pl` スクリプトを連動させる場合は、次の手順で Mascot Server がインストールされている PC に「Swissknife」Perl ライブラリと「`varsplc.pl`」スクリプトをインストールしてください。

1. Swissknife のインストール

EBI の FTP サイト、

<ftp://ftp.ebi.ac.uk/pub/software/swissprot/Swissknife/>

から「`Swissknife_1.67.tar.gz`」(2009/9/25 時点のリリース番号) をダウンロードし、「`C:\temp`」フォルダに保存してください。このファイルを解凍すると、`varsplc.pl` スクリプトが必要とする Perl ライブラリを含む、

`C:\temp\Swissknife_1.67\lib\SWISS`

フォルダが作成されますので、このフォルダを Perl のライブラリフォルダ (通常は `C:\perl\lib`) にコピーしてください (`C:\perl\lib\SWISS`)。

2. varsplc.pl のインストール

EBI の FTP サイト、

<ftp://ftp.ebi.ac.uk/pub/software/swissprot/varsplc/>

から「`varsplc.pl`」をダウンロードし、「`C:\inetpub\mascot\bin`」フォルダに保存してください。

3. SwissProt_varsplic セットアップ手順

この章では、db_update.pl スクリプトを利用して、最新の「Swiss Prot オリジナルエン트리」+「splice isoforms」+「splice variants」+「splice conflicts」で構成された配列データベース「SwissProt_varsplic」を新規にセットアップする手順について説明しています。

なお、必要に応じて次の書類を参照してください。

http://www.matrixscience.jp/pdf/jap_database_setup.pdf

(1) SwissProt_varsplic 用フォルダの作成

次の4つのフォルダを作成してください。

```
C:\inetpub\mascot\sequence\SwissProt_varsplic
C:\inetpub\mascot\sequence\SwissProt_varsplic\current
C:\inetpub\mascot\sequence\SwissProt_varsplic\incoming
C:\inetpub\mascot\sequence\SwissProt_varsplic\old
```

(2) Database Definitions 設定

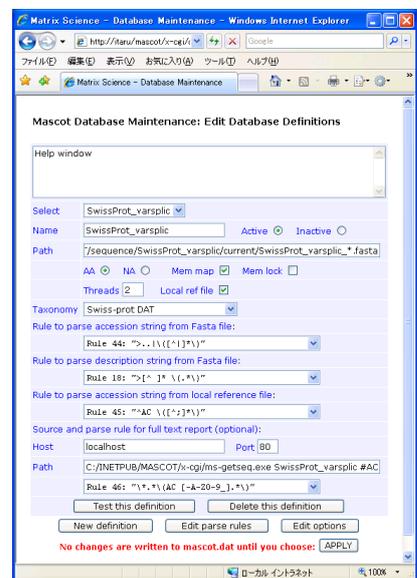
右図のように設定してください。

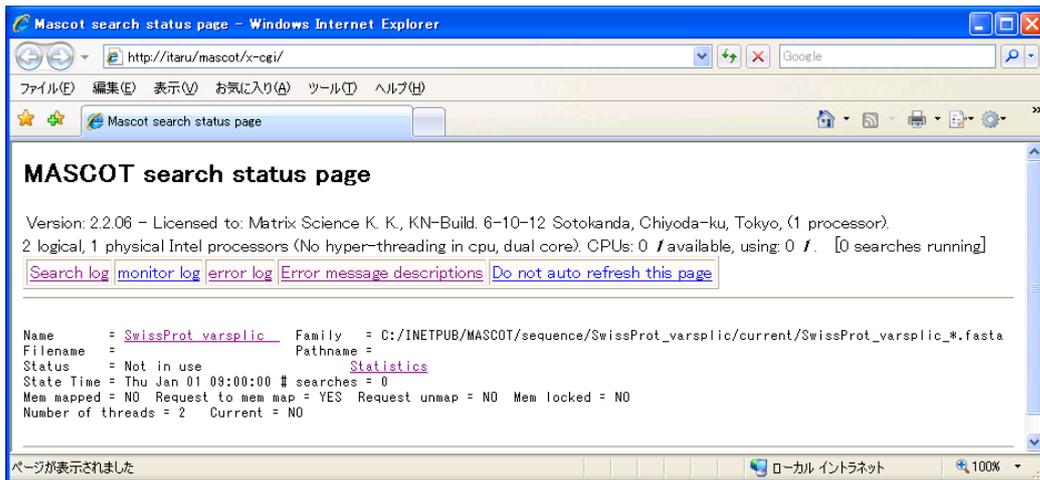
varsplic.pl スクリプトはSwiss Prot の ID ではなくアクセス番号 (AC) を認識しますので、FASTA ファイルおよび DAT ファイルから AC を抽出する設定になっていることに注意してください。

主要な設定項目に対する設定内容は次のようになります。

- [Name] : **SwissProt_varsplic**
- [Taxonomy] : **Swiss-prot DAT**
- [Rule to parse accession string from Fasta file] :
`">.. |%(^[^;]*)"`
- [Rule to parse description string from Fastra file] :
`">[^]* %(.*%)"`
- [Rule to parse accession string from local reference file] :
`"^AC %(^[^;]*)"`
- [Source and parse rule for full text report(optional)] :
 [Host] : **locaohost**
 [Port] : **80**
 [Path] : **"%*. *(AC [-A-Z0-9_]. *)"**

[APPLY] ボタンを押して「SwissProt_varsplic」を有効にしてください。この時点では FASTA ファイル等は存在しませんので、[MASCOT search status page] ページの「SwissProt_varsplic」ブロックは次のように表示されます。





(3) db_update.pl スクリプトの編集

db_update.pl ファイルをテキストエディタで開き、次の<編集前>に「SwissProt_varsplic」を含む3行を追加し、<編集後>のように編集してください(追加文字を赤で示しています)。

<編集前>

```
#
# SwissProt_varsplic_from_EBI
#
} elsif ($ARGV[0] eq "SwissProt_varsplic_from_EBI") {
    $db_name          = "SwissProt";
    $local_incoming_directory = "$SEQUENCE/SwissProt/incoming";
    $local_current_directory = "$SEQUENCE/SwissProt/current";
```

<編集後>

```
#
# SwissProt_varsplic_from_EBI
#
} elsif ($ARGV[0] eq "SwissProt_varsplic_from_EBI") {
#   $db_name          = "SwissProt";
#   $db_name          = "SwissProt_varsplic";
#   $local_incoming_directory = "$SEQUENCE/SwissProt/incoming";
#   $local_incoming_directory = "$SEQUENCE/SwissProt_varsplic/incoming";
#   $local_current_directory = "$SEQUENCE/SwissProt/current";
#   $local_current_directory = "$SEQUENCE/SwissProt_varsplic/current";
```

(4) db_update.pl スクリプトの実行

コマンドプロンプトを起動し、次のコマンドを実行してください。

```
cd c:\inetpub\mascot\bin
db_update.pl SwissProt_varsplc_from_EBI
```

オリジナルの Swiss Prot 関連ファイル (DAT ファイル、Taxonomy 関連ファイル) をダウンロードした後、varsplc.pl スクリプトが実行され、

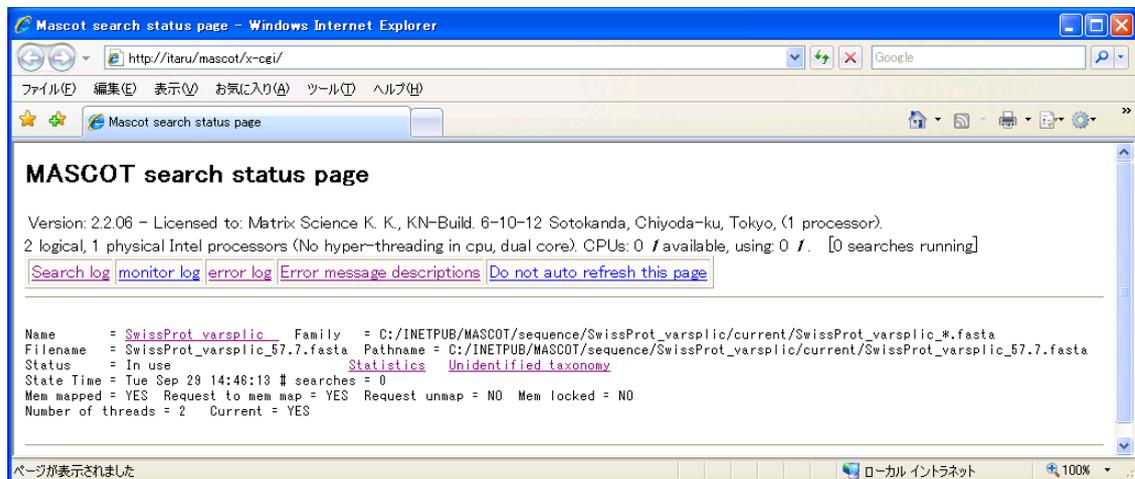
「Swiss Prot オリジナルエントリ」 + 「splice isoforms」 + 「splice variants」 + 「splice conflicts」

で構成された "SwissProt_varsplc" がセットアップされます。なお、すべてのジョブが完了するまでに数時間を要します。

(5) 確認

[MASCOT search status page] ページを開き、"SwissProt_varsplc" のセットアップ状況を確認してください。

[Status] 行が "In use" になればセットアップ完了です。



リリース 57.7 の "SwissProt" のタンパク質エントリ数は 497,293 件であるのに対し、varsplc.pl スクリプトによって作成された "SwissProt_varsplc" のタンパク質エントリ数は 925,162 件となり、タンパク質エントリ数は 2 倍弱に増加しています。

(6) 自動更新

Windows のタスクスケジューラを利用して、定期的に "db_update.pl SwissProt_varsplc_from_EBI" を実行することができます。詳しくは、

http://www.matrixscience.jp/pdf/jap_database_setup.pdf

の「2-4 Windows のタスクスケジューラの設定方法」を参照してください。

4. varsplic 手動実行

Swiss Prot のリファレンスファイル (DAT ファイル) に対して、コマンドプロンプトから直接 varsplic.pl を実行して、スプライシング内容を含む配列データベース (FASTA ファイル、DAT ファイル) を作成したい場合は次のヘルプページを参照してください。

http://www.matrixscience.com/help/seq_db_setup_utilities.html#varsplic

以上



マトリックスサイエンス株式会社

電子メール : support-jp@matrixscience.com

電話 : 03-5807-7895

ファクシミリ : 03-5807-7896

住所 : 〒101-0021 東京都千代田区外神田 6-10-12 KN ビル 3 階